A ONE-TO-ONE REPRESENTATION OF GEODESICS ON A SURFACE OF NEGATIVE CURVATURE.

By Harold Marston Morse.

Introduction.

Surfaces of negative curvature and geodesics upon such surfaces have been considered by J. Hadamard in the paper,* cited below. The paper by Hadamard is in two parts, in the first of which he establishes the existence of a very general class of surfaces of negative curvature; the remainder of Hadamard's paper is devoted to considerations of geodesics upon such surfaces.

In the present paper only those geodesics on the given surfaces of negative curvature are considered which, if continued indefinitely in either sense, lie wholly in a finite portion of space. A class of curves is introduced each of which consists of an unending succession of the curve segments by which the given surface, when rendered simply connected, is bounded. It is shown how a curve of this class can be chosen so as to uniquely characterize some geodesic lying wholly in a finite portion of space. Conversely, it is shown that every geodesic lying wholly in a finite portion of space, is uniquely characterized by some curve of the above class. Use is made of this representation of the geodesics considered to prove several fundamental theorems.

The results of this paper and the representation of geodesics obtained here, will be used in a later paper to establish the existence of a class of geodesics called recurrent geodesics of the discontinuous type.† This class of geodesics offers the first proof that has been given in the general theory of dynamical systems, of the existence of recurrent motions of the discontinuous type.

Part I.

The Surface.

§ 1. We will suppose surfaces without singularities in finite space. We will suppose the surface divisible into overlapping regions such that every point of the surface lying in a finite portion of space is contained as an interior point in some one of a finite number of these regions, and such that

* Les surfaces à courbures opposées et leur lignes geodesiques," Liouville, *Journal de Mathématique*, 5 Sér., t. 4, p. 27, 1898.

† G. D. Birkhoff, "Quelques théorèmes sur le mouvement des systèmes dynamiques," *Bull. de la Soc. Math. de France*, Vol. 40, p. 303, 1912.

the Cartesian coördinates x, y, z, of the points of any one of these regions can be expressed in terms of two parameters, u and v, by means of functions with continuous derivatives up to a convenient order, at least the third, and such that

$$\left(\frac{D(xy)}{D(uv)}\right)^2 + \left(\frac{D(xz)}{D(uv)}\right)^2 + \left(\frac{D(yz)}{D(uv)}\right)^2 \neq 0.$$

By a *curve* on the surface we will understand any set of points on the surface in continuous correspondence with the points of an interval on a straight line, including one, both, or neither of its end points.

We will suppose the Gaussian curvature of the surface to be negative at every point, with the possible exception of a finite number of points, at which points the curvature will necessarily be zero. A first result, given by Hadamard in the paper already referred to, is that a surface of negative curvature cannot be contained in any finite portion of space.

§ 2. By a *funnel* of a surface will be meant a portion of a surface topographically equivalent to either one of the two surfaces obtained by cutting an unbounded circular cylinder by a plane perpendicular to its axis. We will consider surfaces of negative curvature whose points outside of a sufficiently large sphere with center at the origin consist of a finite number of funnels. Each of these funnels will be cut off from the rest of the surface along a simple closed curve. These curves will be taken sufficiently remote on the funnels to be entirely distinct from one another.

An unparted hyperboloid of revolution is an example of a surface of negative curvature with two funnels.

From the definition of a funnel, it follows that by a continuous deformation of the closed curve forming the boundary of the funnel, the funnel may be swept out in such a way that every point of the funnel is reached once and only once. Hadamard considers two classes of funnels: those which can be swept out by closed curves which remain less in length than some fixed quantity, and those which do not possess this property. Surfaces with funnels of the first sort are for several reasons of less general interest than those with funnels of the second sort. In the present paper surfaces with funnels only of the second sort will be considered. Hadamard showed that there exist surfaces of negative curvature possessing funnels all of the second sort of any arbitrary number exceeding one, and such that the surface obtained by cutting off these funnels is of an arbitrary genus.

§ 3. We shall consider surfaces which possess at least two funnels of the second sort, and of the surfaces with just two funnels of the second sort we will exclude those surfaces that are topographically equivalent to an unbounded circular cylinder. Hadamard proves that on the surfaces we

are considering there exists one and only one closed geodesic that is deformable into the boundary of a given funnel, and that this geodesic possesses no multiple points, and no points in common with the other closed geodesics that are deformable into the boundaries of the other funnels.

We shall denote these closed geodesics, say v in number, by

$$g_1, g_2, \cdots, g_v. \tag{1}$$

They will form the complete boundary of a part of the surface, contained in a finite part of space. We denote this bounded surface by S. It may be proved as a consequence of the assumptions concerning the surface, made in § 1, that S is of finite connectivity. According to the well known theory, S may be rendered simply connected as follows: We first cut S along a system of curves,

$$h_1, h_2, \cdots, h_{v-1},$$

each of which has one end point on an arbitrarily chosen point, P on g_v, and the other, respectively, on the geodesic,

$$g_1, g_2, \cdots, g_{v-1}$$

with the same subscript; and no two of which have a point other than P in common, and no points other than their end points in common with the geodesics of the set (1). There then results a surface with a single boundary. This surface can accordingly be rendered simply connected by $2p$ curves,

$$c_1, c_2, \cdots, c_{2p},$$

each of which curves can be taken as beginning and ending at P, and having no other points than P in common with any of the other curves or geodesics.

We denote by T, the simply connected piece of surface obtained by cutting S along the above curves. It may readily be proved as a consequence of the assumptions made concerning the representation of the given surface, that T is topographically equivalent to a plane region consisting of the interior and boundary points of a circle.

The Infinitely Leaved Surface M.

§ 4. We suppose an infinite number of distinct copies of T spread out above the surface S, in such a manner that points arising from a given point of S overhang the given point of S. We denote by

$$A_i^+, \qquad i = 1, 2, \cdots, 2p + v - 1.$$

that boundary piece of T which arises from an arbitrarily chosen side of the ith one of the $2p + v - 1$ cuts,

$$c_1, c_2, \cdots, c_{2p}; \qquad h_1, h_2, \cdots, h_{v-1}, \tag{1}$$

which we have made in S, and by A_i^- that boundary piece of T that arises from the opposite side of the same cut.

Consider now a first copy of T, say M_1, and a particular boundary piece of M_1, say A_i^+. We join M_1 to a second copy of T, joining A_i^+ of M_1 to A_i^- of the second copy of T. In the same manner, we join to M_1 at each different piece of the remaining boundary pieces, a different copy of T, joining each two copies of T along two boundary pieces arising from opposite sides of the same cut. We denote the bounded surface so obtained by M_2.

Proceeding in the same manner, we join to each different boundary piece of M_2 a different copy of T, and one not already a part of M_2, obtaining thereby a new surface M_3. From M_3 we form a new surface M_4, as we formed M_3 from M_2, and continue this process indefinitely, obtaining an enumerable infinity of surfaces each of which contains the preceding. We will understand two different copies of T making up M_n, to have a point in common, if and only if we have expressly joined the two copies together at this point.

§ 5. DEFINITION. *By the surface M we will understand that infinitely leaved surface spread over S that contains each of the surfaces M_n, and every point of which is contained in some one of the surfaces M_n.*

The boundaries of M will consist of those boundary pieces, of the copies of T that make M up, that arise from the geodesics,

$$g_1, g_2, \cdots, g_v.$$

We have made the cuts in S so that on each of the boundary pieces,

$$g_1, g_2, \cdots, g_{v-1}, \tag{2}$$

lies one end point of just one cut. The remaining $4p + v - 1$ end points of the cuts lie at a single point on g_v. On M a sufficiently restricted neighborhood of any one of the points on the geodesics (2), at which lies an end point of a cut, is, when severed along this cut, best described as topographically equivalent to the neighborhood of a point on the boundary of a half plane, when that half plane is severed along a ray distinct from the boundary of the half plane, and issuing from the given point. Similarly a sufficiently restricted neighborhood of that point on g_v at which lie the remaining $4p + v - 1$ end points of cuts, is, when severed along these cuts, topographically equivalent to the neighborhood of a point on the boundary of a half plane, when that half plane is severed along $4p + v - 1$ rays, distinct from each other and from the boundary of the half plane, and issuing from the given point. A sufficiently restricted neighborhood of any other point of M, is topographically equivalent to the complete neighborhood of any point in the plane.

We may conclude that *a sufficiently restricted neighborhood of any point on M is a one-leaved copy of the neighborhood on S of the point overhung on S.*

§ 6. It will be convenient to denote by R a plane region consisting of the points on and interior to a circle.

Any single copy of T is topographically equivalent to a region R. From the method of formation of the surfaces M_n of § 4, it follows that any one of the surfaces M_n is topographically equivalent to a region R. Now from the definition of M in § 5, it follows that any set of points on M that lie in only a finite number of copies of T making up M, will be contained in some one of the surfaces M_n forming a part of M. *Hence any set of points that lie in only a finite number of the copies of T that make up M, will be contained in some region of M, topographically equivalent to a region R.*

§ 7. An infinite set of points on M will be said to have a *limit point P*, if there exists an infinite number of points of the set in every neighborhood of P, on M.

We will prove that a set of points every infinite subset of which has at least one limit point on M, cannot have points in more than a finite number of different copies of T making up M.

If there were points of the set in more than a finite number of the copies of T, there would exist a subset of the given set, say Q, containing an infinite number of points of the given set, and such that each different point would lie in a different copy of T. By the hypothesis made concerning the given set, Q has at least one limit point on M, say P. There are accordingly an infinite number of points of Q in every neighborhood of P, and since no two points of Q lie in the same copy of T, there must be an infinite number of copies of T with points in the neighborhood of P. This is impossible; for all the points in a sufficiently restricted neighborhood of a point on M lie on at most $4p + v$ copies of T, that are joined together at that point.

Hence the result required is proved. Combining this result with the conclusion of the preceding section, we have that *a set of points on M every infinite subset of which has at least one limit point on M, can be included in a region of M, made up of a finite number of copies of T, and topographically equivalent to the plane region consisting of the interior and boundary points of a circle.*

Curves on M Overhanging Curves on S.

§ 8. DEFINITION. Let there be given on S a curve k, and on k a point P. Let P' be any point on M, overhanging P. Starting from a particular point of k, say Q, let P trace out k, first in one sense, and then starting again at Q, in the other sense. As P moves continuously on S, let P' move continuously on M, always overhanging P, and starting at both the times that P starts at Q, at the same point, a point, say Q', overhanging Q. The

continuous curve traced out by P' on M, will be said to *correspond* to k on S.

It follows from the result of § 5, that the curve traced out by P', is uniquely determined by k and the choice of the points Q and Q'.

DEFINITION. Let h be a curve on S that is closed. Let h be cut at some point so as to form a curve segment k. Let h' be any curve segment on M corresponding, in the sense of the preceding definition, to k. h and $h_,$ will be said to *correspond.*

Every curve on M obviously corresponds to one and only one curve on S, while a given curve on S will correspond to an infinite number of different curves on M.

§ 9. *If a curve on S is both closed and deformable into a point, it corresponds on M only to curves that are closed.*

Let h be a curve on S that is both closed and deformable into a point. Let Q' and P' be the end points of a curve segment on M corresponding in the sense of the preceding definition, to h. Since h is closed, Q' and P' overhang a single point on h, say Q.

Suppose now that the above lemma were not true, and that Q' and P' were distinct points on M. Let h be continuously deformed into a point. At the same time let the curve segment joining Q' to P', say h', be continuously deformed on M in such a manner that h' corresponds at each stage of the deformation to the curve into which h has been deformed at that stage, while Q' and P' overhang Q. When, at the end of this process, h reduces to a point, h' must also reduce to a point.

On the other hand P' and Q', the end points of h', in their initial positions were supposed to be distinct points. Since P' and Q' move continuously throughout the deformation, there must then be a first position on M, say P'', at which they are coincident. P' and Q', throughout this deformation, overhung one point of S, namely Q. Hence the point P'' will have in every neighborhood on M pairs of points overhanging the same point of S. This, however, is contrary to the result of § 5, from which contradiction we infer the truth of the above lemma.

§ 10. The following lemma serves as the converse of the preceding.

If a curve on M is closed, it corresponds on S to a curve that is both closed and deformable into a point.

Denote by h the given closed curve on M. It follows from § 7 that there exists a region of M, say M', that contains all the points of h as interior points, and is topographically equivalent to a plane region consisting of the interior and boundary points of a circle. On M', h can be deformed into a point, and hence on M.

Denote by F a continuous family of closed curves on M through the mediation of which h may be deformed into a point on M. On S the

closed curve corresponding to h, may be deformed into a point through the mediation of that continuous family of closed curves which correspond to the family F on M.

§ 11. DEFINITION. Let r be any integer, positive, negative, or zero. Let T_r denote a particular copy of T of M. By a *linear set* of copies of T of M, will be understood a region of M consisting of a set of the copies of T in M of the form,

$$\cdots T_{-2}, T_{-1}, T_0, T_1, T_2, \cdots, \tag{1}$$

or of the form of any subset of consecutive symbols of (1), in which any three successive copies of T are three different copies of T of M, and in which any two successive copies of T are copies of T that in M are joined along a common boundary piece. A linear set which has no first or last copy of T, will be termed an *unending* linear set.

A linear set is of the nature of an unclosed ribbon of surface. That a linear set does not form a multiple-covered region of M will now be proved.

(A) *No copy of T of M appears more than once in any linear set.*

Starting with any copy of T, say T^0, of a given linear set, if possible, let T' be the first successor of T^0 that is identical with some copy of T of the given linear set between T^0 and T', say T''. That linear subset of the given linear set that consists of T'' and its successors up to, but not including T', forms a simply covered, connected region of M, in which the two sides of any boundary piece common to two of its successive copies of T, can be joined by a curve that does not cross that boundary piece. This is im. possible, since each boundary piece common to two copies of T of M joins two boundary points of M and divides M into two regions. Cf. § 5 and § 6.

(B) *If T' and T'' are any two copies of T of the copies of T that make up M, there exists one and only one linear set of which T' is the first, and T'' the last member.*

We form a linear set in which the first copy of T is T' and in which the successor of T', say T_1, is that copy of T that adjoins T along that boundary piece of T' that separates T' from that part of the surface M that contains T''. If T_1 is not T'', its successor shall be that copy of T that adjoins T_1 along that boundary piece of T_1 that separates T_1 from that part of the surface in which T'' lies. We continue this process indefinitely, or until T'' appears in the linear set formed.

Join any interior point of T' to an interior point of T'' by a curve, say k, lying wholly on M. Each copy of T determined in the process of the preceding paragraph contains at least a point of k. But from the result of § 7 we infer that there are only a finite number of copies of T of M that contain any point of k. It follows that the process of the preceding paragraph will lead after a finite number of steps to T''.

There is but one linear set of which T' is the first and T'' the last copy of T. For no such linear set can contain as a part of its boundary that boundary piece of T' that separates T' from that part of the surface that contains T''. Hence, in any such linear set T' and T_1 must appear at least once as successive copies of T. But according to the result (A), no copy of T of M appears more than once in any linear set; it follows that T_1 is the successor of T' in any linear set joining T' to T''.

In a similar manner it follows that any linear set joining T' to T'', contains T_2 as the successor to T_1. Continuing this method of proof, it is seen that every linear set joining T' to T'', is identical with the first such linear set that we have formed.

§ 12. We denote by H, the set of all curve segments on M each of which corresponds on the original surface to some one of the segments,

$$c_1, c_2, \cdots, c_{2p}; \qquad h_1, h_2, \cdots, h_{v-1}; \qquad g_1, g_2, \cdots, g_v.$$

Every boundary piece of the copies of T of M, is included in the set H.

DEFINITION. Let r be any integer, positive, negative, or zero. Let k_r be a particular piece of the set H of M. By a *reduced curve* of the set H will be understood a continuous curve composed of a set of pieces of the set H of the form,

$$\cdots, k_{-2}, k_{-1}, k_0, k_1, k_2, \cdots, \tag{1}$$

or of the form of any subset of consecutive symbols of (1), in which no two consecutive pieces are copies of the same piece of the set H, and which contains no pieces corresponding to g_v on the original surface. A reduced curve without end points will be called an *unending* reduced curve.

(A) *No reduced curve can begin and end at the same point.*

Starting with any point, say P, of a given reduced curve, if possible, let P' be the first point following P that is identical with some point of the reduced curve between P and P', say P''. The segment $P'P''$ of the given reduced curve forms a closed curve without multiple points, which closed curve we denote by C.

According to the result of section 7, C can be included in some simply connected region of M consisting of a finite number of copies of T. C, accordingly, forms the boundary of a simply connected region of M, say R, all of whose points are contained in a finite number of copies of T of M. Since C contains no points interior to any copy of T, R contains each copy of T of which it contains any interior points.

Taking the copies of T in the order in which they were added in section 4 to form M, let T' be the last copy of T of R. T', like every other copy of T of M, is joined along a common boundary piece to but one copy of T of

M, say T'', that was added to M prior to T'. R accordingly contains none of the copies of T that are adjoined to T' along a common boundary piece except T''. Hence all of the boundary of T', except the piece common to T' and T'', must form part of the boundary of R, and hence a part of C. In particular, C contains that piece of the boundary of T' that corresponds on the original surface to the boundary geodesic g_v, contrary to the definition of a reduced curve. From this contradiction we infer the truth of the lemma.

The following result is an immediate consequence of the result (A).

(B) *A reduced curve can have no multiple points.*

(C) *There is one and only one reduced curve joining any two given points on pieces of the set H.*

Denote the two given points by P and Q. Among the curves consisting of any continuous succession of pieces of the set H, there obviously exists a variety of curves joining P to Q. If in any such curve, each piece that corresponds on the original surface to g_v, be replaced by the remainder of the boundary of the copy of T containing that piece, there will then result a curve containing no piece corresponding to g_v on the original surface. Of all curves of the latter class joining P to Q, any one that contains the minimum number of pieces of the set H, will be such that no two successive pieces are copies of the same piece of the set H, and will accordingly be a reduced curve.

If possible, let h' and h'' be two different reduced curves joining P to Q. Tracing out h' and h'' respectively, starting from P, let k' and k'' be the first pieces on h' and h'' respectively, that are different. That portion of h' that begins at Q, and ends with k', followed by that portion of h'' that begins with k'', and ends with Q, is a reduced curve with initial and final point at the same point, namely Q. This is contrary to the result (A) of this section, from which contradiction we infer the truth of the lemma (C).

(D) *A reduced curve never returns to a copy of T which it has once left.*

If this assertion be false, there exists some continuous segment of a reduced curve, say k, with end points, say P and Q, belonging to the same copy of T of M, and with no further points in common with that copy of T. A reduced curve cannot begin and end at the same point; hence P and Q are different points. However, let k' be that part of the boundary of the given copy of T that joins P to Q, and does not contain any piece of the set H corresponding to g_v on the original surface. k', taken with k, forms a segment of a reduced curve that may be considered as beginning and ending at P. We conclude that no such segment as k exists, and the lemma is proved.

§ 13. The following is the first of two theorems showing the relation of reduced curves to linear sets.

(*A*) *An unending linear set of copies of T contains one and only one unending reduced curve.*

Let T^0, T' and T'', be any three successive copies of T in the given linear set. There exist two distinct and continuous portions of the boundary of T', of which one may reduce to a point such that each joins some boundary point of T^0 to a boundary point of T'', without containing any other points of T^0 and T''. We associate with T' that one of these portions that does not contain a piece of the set H corresponding to g_v on the original surface. Denote this portion by k. The set of all curve segments such as k, taken in the order in which they arise from the given linear set, will not necessarily form a continuous curve. However by a proper addition of the boundary pieces common to successive copies of T of the given linear set, the resulting curve will be continuous. If these additions are made only when necessary to make the curve continuous, the resulting curve will also be a reduced curve.

If possible, let r be a second reduced curve contained in the given linear set. As proved in the preceding section, a reduced curve never returns to a copy of T which it has once left; it follows that r contains at least a point of each of the pieces of the set H that separate members of the given linear set, and hence contains each of the points or segments k, associated in the first paragraph of this proof with each copy of T of the given linear set. But according to the result (*C*) of section 12, there is but one segment of a reduced curve joining any two points of M. Hence any reduced curve that contains each of the points or curve segments k, associated with each copy of T of the given linear set, is thereby uniquely determined. There is thus but one reduced curve contained in the given linear set.

The preceding result has its converse in the following:

(*B*) *An unending reduced curve is contained in one and only one linear set, which set is an unending linear set.*

Let

$$\cdots, P_{-2}, P_{-1}, P_0, P_1, P_2, \cdots,$$

be a set of points which lie on the given reduced curve in the order of their subscripts, and which are such that any point of the given curve lies between P_{-n} and P_n for n a sufficiently large positive integer, and where P_{-1} and P_1 are supposed to be taken so as not to lie on a common copy of T of M. Let C_n be the segment joining P_{-n} to P_n of the given reduced curve. Let L_n be a linear set containing P_{-n} and P_n, and made up of the smallest possible number of copies of T.

C_n lies wholly in L_n. If this assertion be false, let k be a piece of the set H which forms part of the boundary of L_n, and contains a point at which C_n leaves L_n. The end points of C_n lie on L_n; hence C_n returns

to L_n. But k divides the whole of M into two regions; hence to return to L_n, C_n would have to return to k, which is impossible, since a reduced curve never returns to a copy of T which it has once left.

We will now prove that if r is any positive integer greater than n, the linear set L_r contains L_n. Let h be any one of the boundary pieces common to two adjacent copies of T of L_n. C_n contains points, not on h, in each of the two regions into which M is divided by h, for otherwise one of the two linear sets into which L_n is divided by h, would contain all the points of C_n and be a linear set containing fewer copies of T than does L_n, contrary to the hypothesis concerning L_n. Hence any connected region that contains all the points of C_n, must contain the two copies of T of L_n that are adjoined along h. Hence any connected region that contains all of the points of C_n, must contain each copy of T of L_n. In particular, L_r contains all the points of C_n, and hence contains L_n.

The set of all copies of T of all regions L_n, for all positive integers n, will accordingly constitute a linear set, which we denote by L. Any point of the given curve is contained on every curve C_n for sufficiently large values of n, and hence is contained in L. We will now prove that L is an unending linear set.

As seen earlier in the proof, the two linear sets into which L_n and hence L is divided by any boundary piece h, common to copies of T of L, each contain points of the given reduced curve that are not on h. Since a reduced curve never returns to a copy of T which it has once left, the removal of h divides the given reduced curve into just two continuous portions, which lie respectively in the two linear sets into which L is divided by h. But each of these continuous portions of the given reduced curve will contain an infinite number of different pieces of the set H, so that each of the two linear sets into which L is divided by h, must contain an infinite number of different copies of T. We conclude that L is an unending linear set.

Any second linear set that contains L, would have to contain each of the linear sets L_n, hence every copy of T of L. Since there is but one linear set joining any two copies of T of M, any linear set containing all the copies of T of L, is identical with L.

The two results of this section may be combined in the following fundamental theorem:

THEOREM I. *There is a one to one correspondence between the set of all unending reduced curves, and the set of all unending linear sets of M, in which each reduced curve corresponds to that linear set in which it is contained.*

PART II.

Some General Properties of Geodesics on M.

§ 14. Hadamard proved that there is on S one and only one geodesic joining two given points, and deformable into an arbitrary curve joining the two given points, and that this geodesic is shorter than any other rectifiable curve deformable into the given curve joining the two given points. With the aid of the results of sections 9 and 10, Hadamard's result becomes the following theorem.

THEOREM II. *There exists on M one and only one geodesic joining two given points, and this geodesic is shorter than any other rectifiable curve joining the two given points.*

COROLLARY I. *On M two geodesics can intersect in but one point.*

COROLLARY II. *On M a geodesic can have no multiple points.*

§ 15. On a surface representable in the manner in which the given surface is representable, there exists one and only one geodesic through a given point, and tangent to a given direction.

DEFINITION. A point on the surface, and a direction tangent to the surface will be called an *element,* and will be said to *define that sensed geodesic* that passes through the initial point of the given element, and is such that its positive tangent direction at that point agrees with the direction of the given element.

If u and v are parameters in any representation of a part of the surface, and if θ' is the angle which a given tangent direction makes at the point $(u'v')$ with the positive tangent to the curve $u = u'$, then $(u'v'\theta')$ will represent an element of the given surface. We shall understand by each statement of metric relations between elements, the same statement of metric relations between the points in space of three dimensions, obtained by considering the complex $(u'v'\theta')$ as the Cartesian coördinates of a point.

§ 16. Let G be any geodesic segment lying on the original uncut surface. G is an extremal in the Calculus of Variations problem of minimizing the arc length, from which theory we can readily obtain the following theorem that describes the nature of the variation of G with variation of its initial element.*

Corresponding to any positive constants, e and h, there exists a positive constant d, so small, that if any two elements, with initial points on the bounded surface S, lie within d of each other, and if a second pair of elements lie respectively on the two geodesics defined by the first two elements, and if further the initial points of this second pair of elements lie respectively at a distance, measured along the given geodesics from the geodesics' initial

* Cf. Bolza, *Vorlesungen über Variationsrechnung,* 1909, p. 219.

points, that is the same in both cases and that does not exceed h, the second pair of elements will lie within e of each other.

§ 17. According to Theorem II, § 14, any given geodesic on M is shorter than any other rectifiable curve joining its end points. According to the theory of the Calculus of Variations, this is a case where there exists on the given geodesic on M no point conjugate to a given point on the given geodesic. A particular consequence* of this result, as given in the theory of the Calculus of Variations, is that, if we vary the end points of a given geodesic segment, there exists further geodesic segments joining these end points and varying continuously with the end points, both in position and in length. We have seen that there is no more than one geodesic segment joining any two points on M. Whence we may say that the length of the geodesic segment, joining two points on the surface M, is a single-valued continuous function of the position of these points. In particular we have the result:

There exists a finite upper limit to the lengths of all geodesic segments joining pairs of points lying on a closed set of points on M.

§ 18. The following statement, of importance in the developments that are to follow, may readily be proved as a consequence of the development given by Hadamard.

Corresponding to two arbitrary positive constants, e and d, there exists a positive constant h, so large, that if on M each point of a first geodesic segment G, of length h, lies within a geodesic distance d of some point of a second geodesic segment, then this second geodesic segment has at least one element within an e of that element of G which lies at the mid point of G.

§ 19. *The Network H Replaced by a Network H' of Geodesic Segments.* We denote by H', the set of all geodesic segments on M each of which joins two end points of a piece of the set H, defined in § 12.

We shall prove the following lemma:

Two pieces of the set H' arising from two different pieces of H, can have no points in common other than one end point.

Let P and Q, P' and Q', be respectively the end points of two different pieces of the set H. Denote by h the geodesic segment that joins P to Q, and by h' the geodesic segment that joins P' to Q'. All cases are included in the three following cases:

CASE I. The two given pieces of the set H have an end point in common.

In particular suppose $P = P'$; if $P = P'$, it follows from the nature of the construction of the surface M, that $Q \neq Q'$. Hence, if $P = P'$, the geodesic segments, h and h', have P as a common end point, and are not identical, since $Q \neq Q'$. With the aid of Corollary II, § 14, we conclude that h and h' can have no other point than P in common.

* Cf. Bolza, loc. cit., p. 307.

CASE II. $P \neq P'$, and $Q \neq Q'$, and one or both of the geodesic segments, h and h', form a piece of the geodesic boundary of the surface M.

h and h' are not identical, since $P \neq P'$, and $Q \neq Q'$. In case h were a piece of the geodesic boundary of M, h' could not meet h in other than an end point without passing off from the surface at that point, contrary to the result of Theorem II, § 14.

CASE III. $P \neq P'$, and $Q \neq Q'$, while neither h nor h' form a piece of the geodesic boundary of the surface M.

According to the result of § 7, there exists on M, a simply connected region, bounded by a simple closed curve, and containing the two given pieces of the set H and the geodesic segments h and h'. This region, which we denote by R, we will suppose taken so large that no points of the two given pieces of the set H, nor of the geodesic segments h and h', are boundary points of R unless they are boundary points of M.

P, Q, P', *and* Q', all lie on the boundary of M, and hence on the boundary of R. Since $P \neq P'$, and $Q \neq Q'$, the two given pieces of the set H have no points in common; it follows that P, Q, P', and Q', lie on the boundary of R in the circular order named.

On the other hand, it is an hypothesis of this case, that neither h nor h' are a piece of the geodesic boundary of M; it follows that neither h nor h' have points other than their end points on the boundary of M, and hence of R. A particular consequence is that h divides R into two regions. If now h' crossed h, its end point P' would lie in one of the two regions into which R is divided by h, while its remaining end point, Q', would lie in the other such region. P, P', Q, Q', would thus lie on the boundary of R in the circular order named. From this contradiction we infer the truth of the lemma in this case. The proof is thus given in general.

§ 20. We seek now to render the original surface simply connected by means of cuts made along geodesics.

In accordance with the result given in § 14, the end points of each one of the cuts by means of which the original surface was rendered simply connected, can be joined on the original surface by a geodesic segment deformable into the given cut. With the aid of the result of § 9, it appears that these geodesic segments correspond on M to the geodesic segments of H', and further that two of these geodesic segments arising from different cuts correspond on M only to different members of the set H'. From the result of the preceding section we conclude that no two of these geodesic segments have any points in common other than their end points.

If now the original surface be cut along these geodesic segments, the resulting surface, which we denote by U, will be simply connected. The infinitely leaved surface M can be formed by joining together copies of U in the same manner as it was formed from copies of T.

We define a reduced curve of the set H', and a linear set of copies of U, by replacing H and T in the definitions of sections 11 and 12, by H' and U, respectively. In the same manner, we obtain from the results of sections 11 and 13, the following results in terms of H' and U.

(A) *If U' and U'' are any two copies of U of the copies of U that make up M, there exists one and only one linear set of which U' is the first and U'' the last member.*

(B) *There is a one to one correspondence between the set of all unending reduced curves of the set H', and the set of all unending linear sets of copies of U, in which each reduced curve corresponds to that linear set in which it is contained.*

Geodesics Lying Wholly on M.

§ 21. Let G be a geodesic lying wholly on M. G cannot become infinite in length in any one copy of U, as follows from the result of section 17. In leaving a copy of U of M, G cannot be tangent to any one of the geodesic segments that separate that copy of U from the remainder of M. For in such a case G would coincide with that geodesic segment, and pass off from M at its end points. Further, it follows from Corollary I, section 14, that G can have but one point of intersection with any of the geodesic segments that separate the different copies of U of M. Hence G never returns to a copy of U of M which it has once left. Since G cannot become infinite in length in any one copy of U, it follows that either of the two portions into which G may be divided by an arbitrary one of its points, has points in common with an infinite number of different copies of U.

From these last two results, it follows by a proof, which except for terminology, may be given as a repetition of the proof of (B), section 13, that *a geodesic lying wholly on M, is contained in one and only one linear set, which set must be an unending linear set.*

§ 22. As a converse to the preceding result, we have the following:

If there be given any unending linear set, there exists one, and only one geodesic contained wholly in the given linear set, and this geodesic has at least one point in each copy of U of the given linear set.

Let an unending linear set of copies of U be given as follows:

$$\cdots,\ U_{-2},\ U_{-1},\ U_0,\ U_1,\ U_2,\ \cdots.$$

Let g_n be a geodesic segment joining any interior point of U_{-n} to an interior point of U_n. The set of copies of U which g_n passes through, is seen to form a linear set joining U_{-n} to U_n. But from the result of section 11, it follows that there is but one such linear set joining U_{-n} to U_n. The linear set,

$$U_{-n},\ U_{-n+1},\ \cdots,\ U_0,\ \cdots,\ U_{n-1}U_n,$$

is one such set; we conclude that this is the set of copies of U through which g_n passes.

Denote by E_n an element on that part of g_n that lies in U_0. The set of all elements E_n will have a limit element. Let G be the geodesic defined by this element. We will first show that G has a point in each copy of U of the given linear set.

Let r be any positive integer. For integers $n > r$ that portion of g_n in

$$U_{-r},\ U_{-r+1},\ \cdots,\ U_0,\ \cdots,\ U_{r-1}U_r, \tag{1}$$

is according to the result of section 17, less in length than some fixed quantity independent of n. Now a finite segment of a geodesic varies continuously with its initial element. It follows that G possesses a finite segment, say G_r, which has a point in each copy of U of (1), and which is wholly contained in the set (1). From the fact that G_r has a point in each copy of U of (1), we may conclude that G has a point in each copy of the given linear set.

We seek to prove that G is wholly contained in the given linear set. Any finite segment of any geodesic on M has points in not more than a finite number of copies of U. Cf. section 7. We may conclude that, for r sufficiently large, any given segment of G that begins with a point of U_0, is included in one of the two portions into which G_r is divided by that point. Thus every point of G lies on some segment G_r. But every point of G_r, and hence every point of G lies in the given linear set.

If there were a second geodesic, say G', contained in the same linear set as G, it would follow from the result of section 18, that every element of G would be a limit element of elements on G'. This is impossible, since G' can never return to a copy of U which it has once left.

The results of this section and the preceding, are summed up in the following:

THEOREM III. *There is a one-to-one correspondence between the set of all geodesics lying wholly on M, and the set of all unending linear sets, in which each geodesic corresponds to that linear set in which it is contained.*

A similar theorem is now obtained from (B) in § 20.

THEOREM IV. *There is a one-to-one correspondence between the set of all geodesics lying wholly on M, and the set of all reduced curves of the set H', in which each geodesic corresponds to that reduced curve that is contained in the same linear set.*

Congruent Curves.

§ 23. DEFINITION. Two curves on M that correspond to the same curve on S will be said to be *congruent*.

The first statement of the following theorem serves as an existence proof for closed geodesics on the original surface.

(*A*) *If a reduced curve on M consists of successive mutually congruent portions, the geodesic that lies in the same linear set as the given reduced curve, consists also of successive mutually congruent portions. Conversely, if a geodesic lying wholly on M consists of successive mutually congruent portions, the reduced curve that lies in the same linear set, consists also of successive mutually congruent portions.*

Suppose the given reduced curve consists of successive mutually congruent portions. Let the configuration, consisting of the given reduced curve and the geodesic lying in the same linear set, be carried as a whole into that congruent configuration in which one of the successive mutually congruent portions of the given reduced curve is carried into the succeeding congruent portion. Denote this transformation by T.

The given geodesic is carried into itself by T. For otherwise there would exist two different geodesics, namely the given geodesic and the geodesic into which it is carried by T, both lying in the same linear set as the given reduced curve. This is contrary to the result of Theorem IV, section 22.

Now let P' be any point on the given geodesic. Let P'' be the point in which P' is carried by T. P'' lies on the given geodesic, since the given geodesic is carried into itself by T. The successive images of the geodesic segment $P'P''$ which result through repetition of T and its inverse, will be successive mutually congruent portions of the given geodesic, all of whose points can be reached by a sufficient number of such transformations.

The converse statement of the theorem can be proved in a similar manner.

In the same manner also, the following can be established.

(*B*) *Let there be given a first reduced curve and geodesic contained in a single linear set, and a second reduced curve and geodesic that are also contained in a single linear set. If the two given reduced curves are congruent, the two geodesics must also be congruent; and conversely, if the two given geodesics are congruent, the two reduced curves must also be congruent.*

Variation of Geodesics with Their Initial Elements.

§ 24. The following statement depends upon the result given in section 18.

(*A*) Corresponding to any positive quantity e, there exists a positive integer n, so large, that if any two unending linear sets have in common a region U, say U', together with the first n regions U succeeding U' in either sense in the given linear set, then there exists in U', and on each of the two geodesics that pass respectively through the two given linear sets, at least one element that lies within e of some element on that part of the other geodesic that lies in U'.

The converse statement depends upon the property of continuous variation of a geodesic with its initial element, as given in section 16.

(B) Corresponding to any positive integer n, there exists a positive constant e, so small, that if on each of two geodesics there exists some element within e of some element on the other, then the two linear sets through which the two given geodesics respectively pass, possess in common that linear set that contains the region or regions U, in which the two given elements lie, together with n regions U succeeding these regions in either sense.

By virtue of the relations between unending reduced curves and corresponding unending linear sets, as given in the proofs of section 13, the statements of (A) and (B) of this section become the following:

THEOREM VI. *Corresponding to any positive constant e, there exists a positive constant k, so large, that if two unending reduced curves possess in common a continuous segment of length exceeding k, the two corresponding geodesics each have at least one element within e of some element on the other, and with initial point in the same copy of U as the mid point of the common reduced segment. Conversely, corresponding to any positive constant k, there exists a positive constant d, so small, that if on each of two geodesics there exists some element within d of some element on the other, the two corresponding reduced curves possess in common a segment of length k, with mid point in the same copy of U as the initial point of either of the two elements.*

§ 25. DEFINITION. Two elements E' and E'' of a set of elements E on M, will be said to be *mutually accessible* in E, if corresponding to any positive constant e, there exists in the set E a finite ordered subset of elements of which the first element is E', and the last element E'', while each element of the subset, excepting the last, lies within e of the following element.

THEOREM VI. *In the set of all elements on a closed region of M, and on geodesics lying wholly on M, no two elements lying on different geodesics are mutually accessible.*

Let G' and G'' be two different geodesics lying wholly on M. Let E' and E'' be two elements lying respectively on G' and G''. Let R be any closed region of M, in which the initial points of E' and E'' lie. We will show that E' and E'' are not mutually accessible in the set of all elements on R, and on geodesics lying wholly on M.

G' and G'' are different geodesics, and are accordingly not contained in the same linear set of copies of U. There therefore exists some geodesic segment, of the geodesic segments that separate different copies of U of M, that G' crosses and that G'' does not cross. Denote this geodesic segment by h. Denote by B' the set of all elements on R, and on geodesics that lie

wholly on M and cross h. Denote by B'' the set of all elements on R, and that lie on geodesics that lie wholly on M and do not cross h.

B' is a closed set. For it follows from the result of sections 16 and 17, that a limit element of elements of the set B', defines a geodesic, say G, lying wholly on M, and that is either tangent to h or else crosses h. But G cannot be tangent to h without passing off from M at the end points of h. Hence G crosses h. Thus B' is a closed set. In a similar manner it follows that B'' is a closed set.

From their definitions it follows that B' and B'' can have no elements in common. There accordingly exists a positive constant d, such that no element of B' lies within d of any element of B''. Hence E', which belongs to B', and E'', which belongs to B'', cannot be mutually accessible in the set of elements comprising the elements of B' and B''. But the elements of B' and B'' comprise all the elements on geodesics lying wholly on M, and with initial points in the region R.

The theorem thus is proved.

HARVARD UNIVERSITY,
June, 1917.